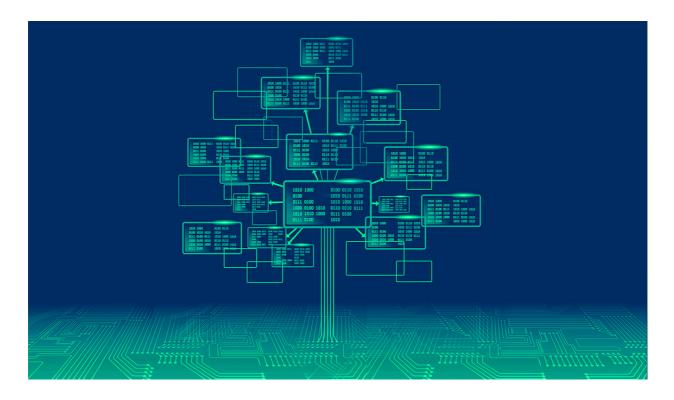
THE XORIANT BLOG

PRODUCT ENGINEERING (HTTPS://WWW.XORIANT.COM/BLOG/CATEGORY/PRODUCT-ENGINEERING)

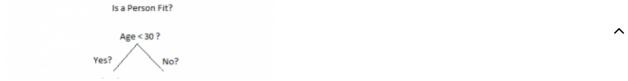
Decision Trees for Classification: A Machine Learning Algorithm

September 7, 2017 by Mayur Kulkarni



Introduction

Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split.



We use cookies on this site to enhance your user experience

Trees-modified-1.png)

An example of a decision tree can be explained using above binary tree. Let's say you want to predict whether a person is fit given their information like age, eating habit, and physical activity, etc. The decision nodes here are questions like 'What's the age?', 'Does he exercise?', 'Does he eat a lot of pizzas'? And the leaves, which are outcomes like either 'fit', or 'unfit'. In this case this was a binary classification problem (a yes no type problem).

There are two main types of Decision Trees:

1. Classification trees (Yes/No types)

What we've seen above is an example of classification tree, where the outcome was a variable like 'fit' or 'unfit'. Here the decision variable is **Categorical**.

2. Regression trees (Continuous data types)

Here the decision or the outcome variable is Continuous, e.g. a number like 123.

Working

Now that we know what a Decision Tree is, we'll see how it works internally. There are many algorithms out there which construct Decision Trees, but one of the best is called as **ID3 Algorithm**. ID3 Stands for **Iterative Dichotomiser 3**.

Before discussing the ID3 algorithm, we'll go through few definitions.

Entropy

Entropy, also called as Shannon Entropy is denoted by H(S) for a finite set S, is the measure of the amount of uncertainty or randomness in data.

$$H(S) = \sum_{x \in X} p(x) \log_2 \frac{1}{p(x)} \frac{1}{(\text{https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-Trees-modified-2.jpg)}$$

Intuitively, it tells us about the predictability of a certain event. Example, consider a coin toss whose probability of heads is 0.5 and probability of tails is 0.5. Here the entropy is the highest possible, since there's no way of determining what the outcome might be. Alternatively, consider a coin which has heads on both the sides, the entropy of such an event can be predicted perfectly since we know beforehand that it'll always be heads. In other words, this event has **no randomness** hence it's entropy is zero.

In particular, lower values imply less uncertainty while higher values imply high uncertainty.

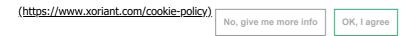
Information Gain

Information gain is also called as Kullback-Leibler divergence denoted by IG(S,A) for a set S is the effective change in entropy after deciding on a particular attribute A. It measures the relative change in entropy with respect to the independent variables.

$$IG(S,A) = H(S) - H(S,A) \underbrace{ (\text{https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-Trees-modified-3.jpg)}_{} (\text{https://www.xoriant.com/blog/wp-content/$$

Alternatively,

We use cookies on this site to enhance your user experience



where IG(S, A) is the information gain by applying feature A. H(S) is the Entropy of the entire set, while the second term calculates the Entropy after applying the feature A, where P(x) is the probability of event x.

Let's understand this with the help of an example

Consider a piece of data collected over the course of 14 days where the features are Outlook, Temperature, Humidity, Wind and the outcome variable is whether Golf was played on the day. Now, our job is to build a predictive model which takes in above 4 parameters and predicts whether Golf will be played on the day. We'll build a decision tree to do that using **ID3 algorithm**.

Day	Outlook	Temperature	Humidity	Wind	Play Golf
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

ID3 Algorithm will perform following tasks recursively

- 1. Create root node for the tree
- 2. If all examples are positive, return leaf node 'positive'
- 3. Else if all examples are negative, return leaf node 'negative'
- 4. Calculate the entropy of current state H(S)
- 5. For each attribute, calculate the entropy with respect to the attribute 'x' denoted by H(S, x)
- 6. Select the attribute which has maximum value of IG(S, x)
- 7. Remove the attribute that offers highest IG from the set of attributes
- 8. Repeat until we run out of all attributes, or the decision tree has all leaf nodes.

Now we'll go ahead and grow the decision tree. The initial step is to calculate H(S), the Entropy of the current state. In the above example, we can see in total there are 5 No's and 9 Yes's.

Yes No Total

We use cookies on this site to enhance your user experience

$$\frac{\text{(https://www.xoriant.com/cookie-policy)}}{\text{No, give me more info}} \qquad \text{OK, I agree}$$

$$Entropy(S) = -\left(\frac{9}{14}\right)\log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right)\log_2\left(\frac{5}{14}\right) \qquad \frac{\text{(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-policy)}}{9}$$

$$= 0.940$$

Trees-modified-5.jpg)

Remember that the Entropy is 0 if all members belong to the same class, and 1 when half of them belong to one class and other half belong to other class that is perfect randomness. Here it's 0.94 which means the distribution is fairly random.

Now the next step is to choose the attribute that gives us highest possible Information Gain which we'll choose as the root node.

Let's start with 'Wind'

$$IG(S,Wind) = H(S) - \sum_{i=0}^{n} P(x) * H(x)$$
 (https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-Trees-

modified-6.jpg)

where 'x' are the possible values for an attribute. Here, attribute 'Wind' takes two possible values in the sample data, hence x = {Weak, Strong}

We'll have to calculate:

- 1. $H(S_{w \sigma \alpha k})$
- 2. H(S_{strong})
- 3. P(Sweak)
- 4. P(Sstrong)
- 5. H(S)=0.94 which we had already calculated in the previous example

(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-Trees-modified-77.jpg)Amongst all the 14 examples we have 8 places where the wind is weak and 6 where the wind is Strong.

$$P(S_{weak}) = \frac{Number\ of\ Weak}{Total}$$

$$= \frac{8}{14}$$

$$P(S_{strong}) = \frac{Number\ of\ Strong}{Total}$$

$$= \frac{6}{14}$$

$$(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-$$

Trees-modified-8.jpg)

Now out of the 8 Weak examples, 6 of them were 'Yes' for Play Golf and 2 of them were 'No' for 'Play Golf'. So, we have,

$$Entropy(S_{weak}) = -\left(\frac{6}{8}\right)\log_2\left(\frac{6}{8}\right) - \left(\frac{2}{8}\right)\log_2\left(\frac{2}{8}\right)$$

(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-

We use cookies on this site to enhance your user experience

'No' for Play Golf.

$$Entropy(S_{strong}) = -\left(\frac{3}{6}\right)\log_2\left(\frac{3}{6}\right) - \left(\frac{3}{6}\right)\log_2\left(\frac{3}{6}\right)$$

$$= 1.000$$
 (https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-

Trees-modified-10.jpg)

Remember, here half items belong to one class while other half belong to other. Hence we have perfect randomness.

Now we have all the pieces required to calculate the Information Gain,

$$IG(S,Wind) = H(S) - \sum_{i=0}^{n} P(x) * H(x)$$

$$IG(S,Wind) = H(S) - P(S_{weak}) * H(S_{weak}) - P(S_{strong}) * H(S_{strong})$$

$$= 0.940 - \left(\frac{8}{14}\right)(0.811) - \left(\frac{6}{14}\right)(1.00)$$

$$= 0.048$$
(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-

Trees-modified-11.jpg)

Which tells us the Information Gain by considering 'Wind' as the feature and give us information gain of **0.048**. Now we must similarly calculate the Information Gain for all the features.

modified-12.jpg)

We can clearly see that IG(S, Outlook) has the highest information gain of 0.246, hence we chose Outlook attribute as the root node. At this point, the decision tree looks like.



modified-13.jpg)

Here we observe that whenever the outlook is Overcast, Play Golf is always 'Yes', it's no coincidence by any chance, the simple tree resulted because of the highest information gain is given by the attribute Outlook.

Now how do we proceed from this point? We can simply apply recursion, you might want to look at the algorithm steps

We use cookies on this site to enhance your user experience

Next step would be computing $H(S_{sunny})$. (https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-content/uploads/2017

Trees-modified-33.jpg)

Table where the value of Outlook is Sunny looks like:

Temperature	Humidity	Wind	Play Golf
Hot	High	Weak	No
Hot	High	Strong	No
Mild	High	Weak	No
Cool	Normal	Weak	Yes
Mild	Normal	Strong	Yes

$$H(S_{sunny}) = \left(\frac{3}{5}\right) \log_2\left(\frac{3}{5}\right) - \left(\frac{2}{5}\right) \log_2\left(\frac{2}{5}\right) = 0.96$$

$$\underline{\text{(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-linearity)}}$$

Trees-modified-14.jpg)

In the similar fashion, we compute the following values

$$IG(S_{sunny}, Humidity) = 0.96$$

$$IG(S_{sunny}, Temperature) = 0.57 \\ \underline{\text{(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-Trees-modified-IG(S_{sunny}, Wind)} = 0.019$$

15.jpg)

As we can see the **highest Information Gain is given by Humidity**. Proceeding in the same way with S_{rain}

(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-Trees-modified-55.jpg) will give us Wind as the one with highest information gain. The final Decision Tree looks something like this.

The final Decision Tree looks something like this.



We use cookies on this site to enhance your user experience

(https://www.xoriant.com/cookie-policy)

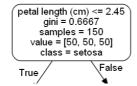
No, give me more info

OK, I agree

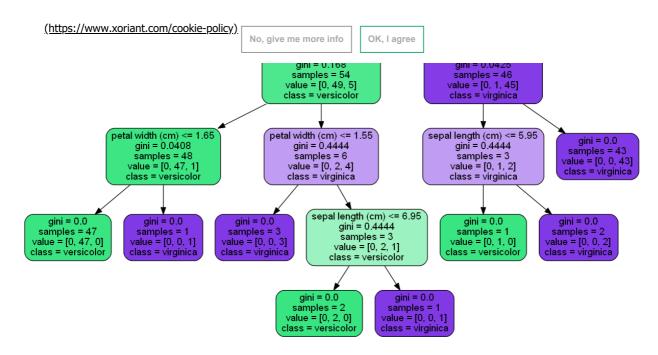
Code:

Let's see an example in Python

```
import pydotplus
from sklearn.datasets import load_iris
from sklearn import tree
from IPython.display import Image, display
__author__ = "Mayur Kulkarni <mayur.kulkarni@xoriant.com>"
 def load_data_set():
    Loads the iris data set
    :return:
                    data set instance
   iris = load iris()
    return iris
 def train_model(iris):
    Train decision tree classifier
                   iris data set instance
    :param iris:
    :return:
                    classifier instance
   clf = tree.DecisionTreeClassifier()
    clf = clf.fit(iris.data, iris.target)
    return clf
 def display_image(clf, iris):
    Displays the decision tree image
    :param clf:
                   classifier instance
    :param iris:
                    iris data set instance
   dot_data = tree.export_graphviz(clf, out_file=None,
                                    feature_names=iris.feature_names,
                                    class names=iris.target names,
                                    filled=True, rounded=True)
    graph = pydotplus.graph from dot data(dot data)
    display(Image(data=graph.create_png()))
 if name == ' main ':
    iris data = load iris()
    decision_tree_classifier = train_model(iris_data)
    display_image(clf=decision_tree_classifier, iris=iris_data)
```



We use cookies on this site to enhance your user experience



(https://www.xoriant.com/blog/wp-content/uploads/2017/08/Decision-Trees-modified-88.png)

Conclusion:

Below is the summary of what we've studied in this blog:

- Entropy to measure discriminatory power of an attribute for classification task. It defines the amount of randomness in attribute for classification task. Entropy is minimal means the attribute appears close to one class and have a good discriminatory power for classification
- 2. Information Gain to rank attribute for filtering at given node in the tree. The ranking is based on high information gain entropy in decreasing order.
- 3. The recursive ID3 algorithm that creates a decision tree.

References:

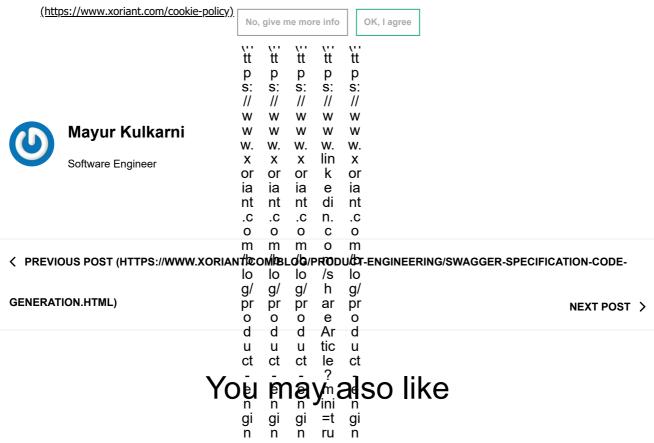
- Information Gain: https://en.wikipedia.org/wiki/Information_gain in decision_trees

 (https://en.wikipedia.org/wiki/Information_gain in decision_trees)
- Entropy: https://en.wikipedia.org/wiki/Entropy_(information_theory))
- 3. ID3: https://en.wikipedia.org/wiki/ID3_algorithm (https://en.wikipedia.org/wiki/ID3_algorithm)
- 4. Entropy YouTube video: https://www.youtube.com/watch?v=O_7lAqni7A (https://www.youtube.com/watch?v=O_7lAqni7A)
- ID3 Example: https://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm)
 (https://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm)

Data Science (https://www.xoriant.com/blog/tag/data-science)

Decision Tree (https://www.xoriant.com/blog/tag/decision-tree)

We use cookies on this site to enhance your user experience





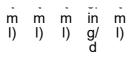
(https://www.xoriant.com/blog/big-data-analytics/aughen@f-analyticseaccelerating-data-insights-into-actions.html)

h h h gi h
Augmented Analytics: Accelerating Data Insights interActions (特許ps://www.pxpriant.com/blog/big-data-analytics/augmented-analytics-accelerating-data-insights-into-actions.html)

h h h gi h
Augmented Analytics: Accelerating Data Insights interActions inte

https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html

March 27, 2019



We use cookies on this site to enhance your user experience

(https://www.xoriant.com/cookie-policy)
No, give me more info
OK, I agree



(https://www.xoriant.com/blog/big-data-analytics/predictive-analytics-part-2.html)

Predictive Analytics – PART 2 (https://www.xoriant.com/blog/big-data-analytics/predictive-analytics-part-2.html)

February 27, 2019



 $\underline{(https://www.xoriant.com/blog/big-data-analytics/predictive-analytics-part-1.html)}$

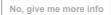
 $Predictive\ Analytics-PART\ 1\ (https://www.xoriant.com/blog/big-data-analytics/predictive-analytics-part-1.html)$

February 20, 2019

11 comments

We use cookies on this site to enhance your user experience

(https://www.xoriant.com/cookie-policy)



OK, I agree



September 12, 2017 at 7:13 am (https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html#comment-665587)

Very intellectual and insightful.!!

REPLY (HTTPS://WWW.XORIANT.COM/BLOG/PRODUCT-ENGINEERING/DECISION-TREES-MACHINE-LEARNING-ALGORITHM.HTML? REPLYTOCOM=665837#RESPOND)

Mayur Kulkarni



September 13, 2017 at 10:50 am (https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html#comment-665837)

Glad that you liked our blog.

REPLY (HTTPS://WWW.XORIANT.COM/BLOG/PRODUCT-ENGINEERING/DECISION-TREES-MACHINE-LEARNING-ALGORITHM.HTML? REPLYTOCOM=665981#RESPOND)

KRISHNA KADABUR



October 3, 2017 at 5:33 pm (https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html#comment-665981)

nice article

REPLY (HTTPS://WWW.XORIANT.COM/BLOG/PRODUCT-ENGINEERING/DECISION-TREES-MACHINE-LEARNING-ALGORITHM.HTML? REPLYTOCOM=669496#RESPOND)

priya



March 17, 2018 at 9:55 am (https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html#comment-669496)

Hi,

I have read your Datascience blog.It's very attractive and impressive.Very useful to me, I like it your blog Thanks for sharing this blog......

Thank you

priya

Pingback: Catboost - What's your data? - Medium - Coding Videos (https://codingvideos.net/catboost-whats-your-data-medium/)

Pingback: An introductory guide to machine learning - Apro Software (https://www.apro.3dize.com/an-introductory-guide-to-machine-\(\sigma \) learning/)

Pindback: Is machine learning entirely based on decision tree? I Physics Forums (https://www.physicsforums.com/threads/is-machine-We use cookies on this site to enhance your user experience

(https://www.xoriant.com/cookie-policy)
No, give me more info
OK, I agree

REPLY (HTTPS://WWW.XORIANT.COM/BLOG/PRODUCT-ENGINEERING/DECISION-TREES-MACHINE-LEARNING-ALGORITHM.HTML? REPLYTOCOM=696590#RESPOND)

Suhaimi Ariffin



April 16, 2019 at 12:24 pm (https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html#comment-696590)

I like how you use math to model a decision tree. Thanks

REPLY (HTTPS://WWW.XORIANT.COM/BLOG/PRODUCT-ENGINEERING/DECISION-TREES-MACHINE-LEARNING-ALGORITHM.HTML? REPLYTOCOM=697648#RESPOND)

priya



May 8, 2019 at 1:50 pm (https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html#comment-697648)

Good Post! Thank you so much for sharing this pretty post, it was so good to read and useful to improve my knowledge as updated one, keep blogging.

Leave a Reply

Your email address will not be published. Required fields are marked *				
Name*				
Email*				

Vebsite	
	^
We	e use cookies on this site to enhance your user experience
<u>(htt</u>	ps://www.xoriant.com/cookie-policy) No, give me more info OK, I agree
POST	COMMENT
SUBSCF	RIBE TO OUR BLOG
lame *	
mail *	
By s	ubmitting your information to Xoriant, you agree to our privacy (/privacy-policy) and cookie (/cookie-policy) policies.
	SUBSCRIBE
Looking fo	or something?
	SEARCH
CATEGO	DRIES

Agile (https://www.xoriant.com/blog/category/agile) (7)

Big Data and Analytics (https://www.xoriant.com/blog/category/big-data-analytics) (15)

Cloud and Infrastructure (https://www.xoriant.com/blog/category/cloud-infrastructure) (4)

Digital (https://www.xoriant.com/blog/category/digital) (1)

Internet of Things (https://www.xoriant.com/blog/category/internet-of-things) (2)

Mobility (https://www.xoriant.com/blog/category/mobility) (23)

Product Engineering (https://www.xoriant.com/blog/category/product-engineering) (47)

TAGS

.NET (https://www.xoriant.com/blog/tag/net)

Adobe (https://www.xoriant.com/blog/tag/adobe)

Agile (https://www.xoriant.com/blog/tag/agile)

We use cookies on this site to enhance your user experience

(https://www.xoriant.com/cookie-policy)
No, give me more info
OK, I agree

Application Development (https://www.xoriant.com/blog/tag/application-development)

automation (https://www.xoriant.com/blog/tag/automation)

Automation Testing (https://www.xoriant.com/blog/tag/automation-testing)

BI (https://www.xoriant.com/blog/tag/bi)

Big Data (https://www.xoriant.com/blog/tag/big-data)

business intelligence (https://www.xoriant.com/blog/tag/business-intelligence)

Chatbots (https://www.xoriant.com/blog/tag/chatbots)

Cloud (https://www.xoriant.com/blog/tag/cloud)

cloud computing (https://www.xoriant.com/blog/tag/cloud-computing)

data analytics (https://www.xoriant.com/blog/tag/data-analytics)

Data Science (https://www.xoriant.com/blog/tag/data-science)

Google (https://www.xoriant.com/blog/tag/google)

Google Cloud Platform (https://www.xoriant.com/blog/tag/google-cloud-platform)

HTML (https://www.xoriant.com/blog/tag/html)

HTML5 (https://www.xoriant.com/blog/tag/html5)

Internet of things (https://www.xoriant.com/blog/tag/internet-of-things-2)

IOT (https://www.xoriant.com/blog/tag/iot)

Java (https://www.xoriant.com/blog/tag/java)

JavaScript (https://www.xoriant.com/blog/tag/javascript)

javascript frameworks (https://www.xoriant.com/blog/tag/javascript-frameworks)

machine learning (https://www.xoriant.com/blog/tag/machine-learning)

microservices (https://www.xoriant.com/blog/tag/microservices)

Mobile application development (https://www.xoriant.com/blog/tag/mobile-application-development-2)

MySQL (https://www.xoriant.com/blog/tag/mysql)

Open Source (https://www.xoriant.com/blog/tag/open-source)

Product Engineering (https://www.xoriant.com/blog/tag/product-engineering)

Product Modernization (https://www.xoriant.com/blog/tag/product-modernization)

QA (https://www.xoriant.com/blog/tag/qa)

R (https://www.xoriant.com/blog/tag/r)

software development (https://www.xoriant.com/blog/tag/software-development)

Software Testing (https://www.xoriant.com/blog/tag/software-testing)

Swagger (https://www.xoriant.com/blog/tag/swagger)

We use cookies on this site to enhance your user experience

(https://www.xoriant.com/cookie-policy)

No, give me more info

OK, I agree

Web Application Development (https://www.xoriant.com/blog/tag/web-application-development)

Web Development (https://www.xoriant.com/blog/tag/web-development)

Windows (https://www.xoriant.com/blog/tag/windows)

1248 Reamwood Avenue, Sunnyvale, CA 94089 | Tel:+1 408 743 4400 (tel:+14087434400) |

info@xoriant.com (mailto:info@xoriant.com)

(http://

www.y

outube

.com/c

hannel

/UCU0 (http://

(http://l (https:/ c158x www.sl

(http://f inkedi /plus.g CKKX idesha acebo (http://t n.com/ oogle. COqE re.net/

ok.co witter.c compa com/+ Fwi7R xoriant

m/Xori om/xor nies/1 xoriant 3A/fee corpor

ant) iant) 66996) /posts) d) ation)

© 2018 Xoriant Corporation. All Rights Reserved. | Terms of Use (/terms-use) | Privacy Policy (/privacy-policy) | Cookie Policy (/cookie-policy) | Sitemap (/site-map)